J. Mol. Biol. (1982) **155**, 447–466

Theoretical Aspects of Specific and Non-specific Equilibrium Binding of Proteins to DNA as Studied by the Nitrocellulose Filter Binding Assay

Co-operative and Non-co-operative Binding to a One-dimensional Lattice

G. Marius Clore, Angela M. Gronenborn

Division of Molecular Pharmacology
National Institute for Medical Research
Mill Hill, London NW7 1AA, England

AND

R. Wayne Davies

Department of Biochemistry
University of Manchester
Institute of Science and Technology
Manchester M60 1QD, England

(Received 1 September 1981, and in revised form 16 November 1981)

The analysis of equilibrium binding isotherms obtained by methods such as the nitrocellulose filter binding assay, which measure the fraction, θ, of DNA to which at least one protein molecule is bound, as a function of the free protein concentration (L_F) require a different type of theoretical framework from that required for analysis of conventional equilibrium binding data, in which the number of moles of protein bound per mole of DNA, θ_C, is measured as a function of L_F. The theoretical framework required to analyse equilibrium binding data generated by measuring $\theta(L_F)$ is developed for co-operative and non-co-operative binding of a protein to a large number of non-specific sites and to a specific sites(s) in the presence of a large number of non-specific sites on a DNA molecule. The theory is simple to apply, equations for $\theta(L_F)$ being easy to derive and evaluate, and is suitable for least-squares analysis. Two examples of the application of the theory to the analysis of experimental data are provided for the specific and non-specific binding of the *Eco*RI restriction endonuclease to bacteriophage λ DNA, and for the specific and non-specific binding of the enzyme dihydrofolate reductase from *Lactobacillus casei* to pBR322 and pWDLcB1 DNA, the latter differing from the former only in a $2{\cdot}9 \times 10^3$ base-pair insert containing the *L. casei* dihydrofolate reductase structural gene. The theoretical and experimental advantages and disadvantages of measuring $\theta(L_F)$ rather than $\theta_C(L_F)$ are discussed.

1. Introduction

DNA–protein interactions play a central role in the regulation of gene expression. Thus, an understanding of the mechanisms involved in gene regulation will require

0022–2836/82/080447–20 $02.00/0

a detailed knowledge of the equilibrium, kinetic and structural aspects of the interactions of DNA with DNA-binding proteins. As a prelude to detailed molecular investigations of a particular DNA–protein interaction, it is usual to examine the equilibrium aspects of this interaction. In general, protein–DNA interactions may be classified as "specific" and "non-specific" (von Hippel & McGhee, 1972), based on the difference in affinity of the protein for one or a few sites on a DNA molecule relative to the affinity of the protein for any site on the DNA. A number of proteins, such as histones, basic polypeptides and all melting proteins, bind only non-specifically and show little or no base sequence preference. However, there also exists a group of proteins such as the *lac* and λ repressors, *cro* protein and most of the restriction endonucleases which, in addition to binding non-specifically to DNA, show varying degrees of affinity for particular base sequences.

In the case where only non-specific binding has been examined, most studies reported in the literature have involved measuring the number of moles of protein bound per mole of DNA, θ_C, as a function of the free protein concentration. This is the standard approach for studying the equilibrium binding of a ligand to a macromolecule (Cantor & Schimmel, 1980). However, in the case of DNA–protein interactions, this approach is complicated by the problem of overlapping sites on the DNA (that is to say, the protein covers more than 1 base-pair on the DNA). Nevertheless, appropriate theoretical frameworks have been developed to deal with the co-operative and non-co-operative binding of a ligand (e.g. a protein) to a homogeneous one-dimensional lattice (McGhee & von Hippel, 1974; Schellman, 1974; Schwarz, 1977; Epstein, 1978). These approaches can be extended with relative ease in the non-co-operative case to the introduction of certain limited degrees of heterogeneity, such as the existence of one or a few specific sites. However, in the presence of co-operativity, the equations become extremely involved, and analysis soon becomes intractable.

In contrast to the studies of non-specific binding, most studies on the binding of proteins to specific sites on DNA have involved measuring the fraction, θ, of DNA to which at least one protein molecule is bound as a function of the free protein concentration. The main experimental method that has been used is the nitrocellulose filter binding assay in which DNA, to which at least one protein molecule is bound, is retained by the filter whereas free DNA passes through it (Riggs *et al.*, 1970; Hinkle & Chamberlin, 1972). Another method that has been developed recently is the immunoprecipitation assay (McKay, 1981) in which DNA, to which at least one protein molecule is bound, is separated from free DNA by immunoprecipitation with an antibody raised against the protein. Both methods involve labelling the DNA radioactively. A consequence of measuring θ instead of θ_C as a function of the free protein concentration (L_F) is that the standard equations in the literature for $\theta_C(L_F)$ (McGhee & von Hippel, 1974; Poland, 1978; Cantor & Schimmel, 1980) are inappropriate for the analysis of the equilibrium binding isotherms generated by measuring $\theta(L_F)$, except in the special case where there exists only a *single site* on the DNA molecule to which the protein can bind or where such an approximation is valid (as, for example, in the case of the *lac* repressor binding to *lac* operator; Riggs *et al.*, 1970). Although some attempts to deal with

this problem have been made (Hinkle & Chamberlin, 1972; Giacomini, 1976,1979; Bailey, 1979; Strauss *et al.*, 1980), these have been incomplete as they are only valid in the absence of co-operative binding and in the absence of any significant heterogeneity, such as a single specific site of high affinity in the presence of a large number of non-specific sites. Consequently, the use of such treatments is inappropriate for the analysis of the vast majority of experimental equilibrium filter binding data, a point that many experimental workers seem to be unaware of.

In this paper, we develop the general theoretical framework required for the direct analysis of equilibrium binding isotherms generated by measuring $\theta(L_F)$ for the co-operative and non-co-operative binding of a protein to a large number of non-specific sites and to a specific site(s) in the presence of a large number of non-specific sites on the DNA. In all cases, the equations for $\theta(L_F)$ are easy to derive and evaluate, and are suitable for least-squares analysis of experimental data. The application of the theory is illustrated by two examples of the analysis of experimental equilibrium filter binding data: the first for the specific and non-specific binding of the *Eco*RI restriction endonuclease to bacteriophage λ DNA, and the second for the specific and non-specific binding of the enzyme dihydrofolate reductase from *Lactobacillus casei* to pBR322 and pWDLcB1 DNA, the latter differing from the former only in a 2·9 kb† insert containing the *L. casei* dihydrofolate reductase structural gene.

2. Theory of Equilibrium Filter Binding

As applied to DNA–protein interactions, the experimental parameter monitored by the nitrocellulose filter binding assay and the immunoprecipitation assay is the fraction, θ, of DNA to which at least one protein molecule is bound, defined by:

$$\begin{aligned}\theta &= [\mathrm{DNA_{bound}}]/[\mathrm{DNA_{total}}] \\ &= (Z-1)/Z, \end{aligned} \tag{1}$$

where Z is the binding polynomial as defined by Wyman (1967); i.e. the sum over all the states of DNA binding to protein. The expression for θ (eqn (1)) contrasts to the expression for the number of moles of protein bound per mole of DNA, θ_C, which is the parameter conventionally measured in equilibrium binding studies using other physico-chemical techniques, and given by:

$$\begin{aligned}\theta_C &= [\mathrm{L_{bound}}]/[\mathrm{DNA_{total}}] \\ &= \frac{[\mathrm{L_{free}}]}{Z}\frac{\partial Z}{\partial[\mathrm{L_{free}}]} = \frac{\partial \ln Z}{\partial \ln [\mathrm{L_{free}}]}. \end{aligned} \tag{2}$$

where the ligand L is the protein.

In the following sections, expressions for the binding polynomial Z, derived on the basis of the standard probability theory, are given for those cases of greatest general interest in the field of protein–DNA interactions: namely, co-operative and

† Abbreviations used: kb, 10^3 bases or base-pairs where appropriate; DHFRase, dihydrofolate reductase.

non-co-operative equilibrium binding of a protein to a large number of non-specific sites, and to a specific site(s) in the presence of a large number of non-specific sites.

(a) *Non-specific binding of a protein to DNA*

If there are N independent non-overlapping protein binding sites of equal intrinsic affinity on the DNA, Z will be given by:

$$Z = \sum_{i=0}^{i=N} \binom{N}{i} K^i L_F^i = (1 + KL_F)^N. \tag{3}$$

where L_F is the free protein concentration, K the intrinsic association constant for the binding of the protein to one site on the DNA, and $\binom{N}{i}$ is the binomial coefficient given by $N!/[(N-i)!i!]$. Thus, θ will be given by:

$$\theta = [(1 + KL_F)^N - 1]/(1 + KL_F)^N. \tag{4}$$

This expression for θ is *not* a rectangular hyperbola, in contrast to the corresponding expression for θ_C, which from equations (2) and (3) is simply given by:

$$\theta_C = NKL_F/(1 + KL_F). \tag{5}$$

The difference between the behaviour of θ and θ_C/N as a function of free ligand concentration is illustrated in Figure 1. The curves for θ_C/N are rectangular hyperbolas and, on a logarithmic scale for L_F, exhibit the characteristic symmetry centred about the point $\theta_C/N = 0{\cdot}5$. The concentration of free ligand L_{50}, for which $\theta_C/N = 0{\cdot}5$, is independent of N and, from equation (5), is given by $1/K$. The curves for θ, on the other hand, are clearly asymmetric about the point $\theta = 0{\cdot}5$, and the value of L_{50} is dependent upon N which, from equation (4), is given by:

$$L_{50} = (\sqrt[N]{2} - 1)/K. \tag{6}$$

In the case of non-specific binding of a protein to DNA, however, the binding site for the protein on the DNA will occupy l base-pairs, and each base-pair can act as the start of a binding site (McGhee & von Hippel, 1974). Thus, initially when no protein molecule is bound there are $(N-l+1)$ binding sites available, where N is the total number of base-pairs. However, the maximum number of protein molecules that can bind to DNA is not $(N-l+1)$ but N/l. Thus, Z is actually given by:

$$Z = 1 + \sum_{i=1}^{i=N/l} \left(\prod_{j=1}^{j=i} \left[\frac{N - jl + 1}{j} \right] \right) K^i L_F^i. \tag{7}$$

Equation (7) neglects the further complication that the average number of binding sites available to bind say $i = m$ protein molecules will be slightly less than $(N-ml+1)$, as there will exist states where the number of binding sites available to m protein molecules is less than $(N-ml+1)$ if two or more protein molecules are

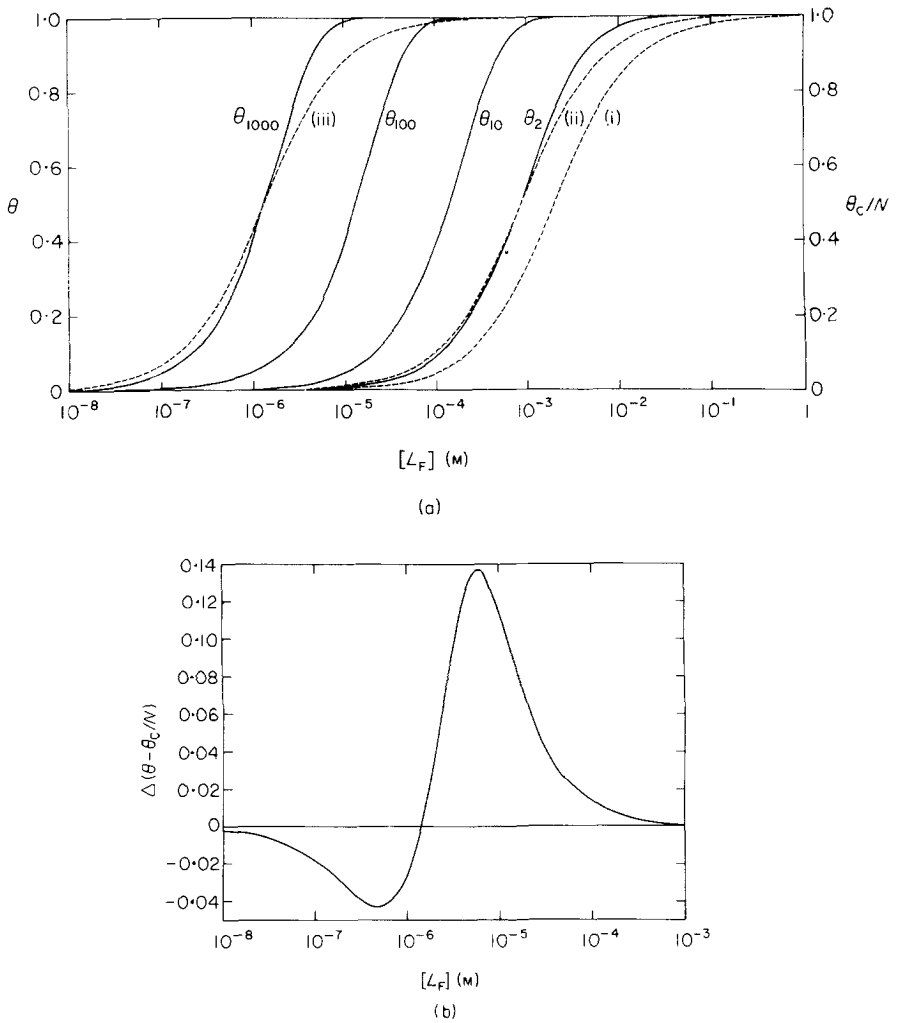

FIG. 1. Illustration of the difference in behaviour of θ (given by eqn (4)) and θ_C/N (given by eqn (5)) as a function of the free ligand concentration (L_F) for the case of N independent non-overlapping sites of equal intrinsic affinity.

In (a) curves for θ are calculated for $K = 500\ \text{M}^{-1}$ and $N = 2$, 10, 100 and 1000; these are shown as continuous lines and are labelled θ_2, θ_{10}, θ_{100} and θ_{1000}, respectively. Curves for θ_C/N are also shown in (a) for $K = 500\ \text{M}^{-1}$, $1207\ \text{M}^{-1}$ and $7{\cdot}21 \times 10^5\ \text{M}^{-1}$; these are represented as broken lines and are labelled (i), (ii) and (iii), respectively. The curves for θ_C/N are rectangular hyperbolas and exhibit the characteristic symmetry centred about the point $\theta_C/N = 0{\cdot}5$. The curves for θ, on the other hand, are clearly not rectangular hyperbolas, and are asymmetric about the point $\theta = 0{\cdot}5$. This is clearly illustrated in (b), where the curve for θ_C/N calculated for $K = 7{\cdot}21 \times 10^5\ \text{M}^{-1}$ is subtracted from the curve for θ calculated for $K = 500\ \text{M}^{-1}$ and $N = 1000$ (for these 2 curves the ligand concentration at which $\theta = \theta_C/N = 0{\cdot}5$ is identical and has a value of $1{\cdot}387 \times 10^{-6}$ M).

separated by a distance of less than l base-pairs (McGhee & von Hippel, 1974; Schellman, 1974; Schwarz, 1977; Epstein, 1978).

However, if $N > 1000$ and $N \gg l$, which will usually be the case experimentally, the expression for Z given by equation (3) in the case of non-overlapping sites and equation (7) in the case of overlapping sites both simplify to:

$$Z = 1 + \sum_{i=1}^{i=5} \frac{N^i K^i L_F{}^i}{i!}. \tag{8}$$

when truncated to degree 5 and making appropriate simplification for large values of N ($N > 1000$), which is sufficiently accurate for the analysis of equilibrium filter binding data† (note N is taken as the total number of base-pairs per DNA molecule).

It is clear that the affinity of the non-specific sites for a protein is unlikely to be homogeneous; rather it will be heterogeneous with a range of affinities. However, providing the range of affinities lies in the range

$$K_i < \sum_{j \neq i}^{N} K_j/10,$$

Z will still be given by equation (8), taking K as the average association constant for the binding of a protein to any non-specific site, and the shape of the equilibrium filter binding curve will be identical to one where the affinity of the non-specific sites for a protein is homogeneous.

Non-specific binding of a protein to DNA is frequently co-operative. To account for co-operativity, we define a co-operativity parameter α for interaction between adjacent sites in terms of a local co-operativity parameter β operating over a distance of m base-pairs from the occupied site such that:

$$\alpha = m(\beta - 1). \tag{9}$$

Z, truncated to degree 5, is then given by:

$$Z = 1 + \sum_{i=1}^{i=5} \frac{\left(\prod_{j=1}^{j=5} [2(j-1)\alpha + N]\right) K_N^i L_F^i}{i!}. \tag{10}$$

The factor $2(j-1)$ multiplying α in equation (10) arises because each occupied site has two sets of adjacent sites, one set on either side. If binding is non-co-operative, $\alpha = 0$ and $\beta = 1$; if binding is positively co-operative, $\alpha > 0$ and $\beta > 1$; if binding is negatively co-operative, $-j < \alpha < 0$ and $0 < \beta < 1$. From equation (10) it is clear that the effect of co-operativity on the shape of the equilibrium filter binding curve will become apparent only when $2\alpha > N/10$. This is illustrated in Figure 2.

† The corresponding polynomial for $\theta(L_F)$ is finite and convergent. By numerical computation, we find that the ratio $r(L_F) = \theta_m(L_F)/\theta_{m+i}(L_F)$, where m and $m+i$ are the number of terms of the polynomial for Z included in the computation of $\theta(L_F)$, lies in the range $1{\cdot}01 > r > 0{\cdot}99$ for all values of L_F when $m \geq 5$, showing that truncation of the polynomial for Z to degree 5 is sufficiently accurate for the analysis of equilibrium filter binding data.

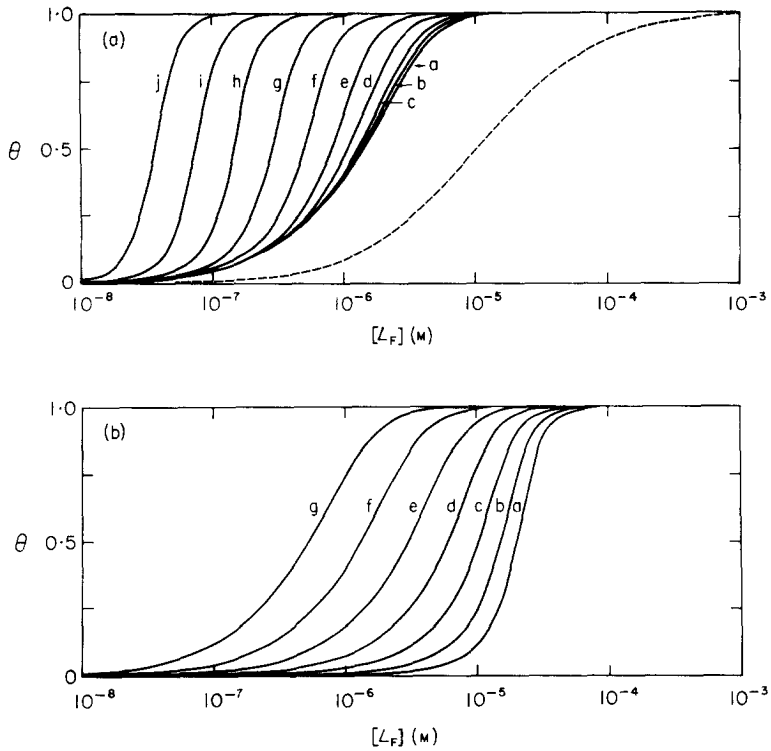

FIG. 2. The effect of co-operativity on the shape of θ as a function of the free ligand concentration (L_F) for the case of N non-specific sites of equal intrinsic affinity.

The curves for $\theta(L_F)$ are calculated using eqns (1) and (10). In (a) the effect on $\theta(L_F)$ of different values of the co-operativity parameter α at constant values of K and N are shown for $K = 500\ \text{M}^{-1}$, $N = 1000$, and $\alpha = 0$, 10 (curve a), 63·1 (curve b), 158 (curve c), 398 (curve d), 10^3 (curve e), $2{\cdot}51 \times 10^3$ (curve f), $6{\cdot}31 \times 10^3$ (curve g), $1{\cdot}58 \times 10^4$ (curve h), $3{\cdot}98 \times 10^4$ (curve i), and 10^5 (curve j). In (b) the effect on $\theta(L_F)$ of different values of N at constant values of K and α are shown for $K = 500\ \text{M}^{-1}$, $\alpha = 10^4$, and $N = 10^3$ (curve a), $2{\cdot}51 \times 10^3$ (curve b), $6{\cdot}31 \times 10^3$ (curve c), $1{\cdot}58 \times 10^4$ (curve d), $3{\cdot}98 \times 10^4$ (curve e), 10^5 (curve f) and $2{\cdot}51 \times 10^5$ (curve g). In (a) a curve for θ_C/N calculated using eqn (5) for $K = 10^5\ \text{M}^{-1}$ is shown for comparison (broken line). Looking at the curves for $\theta(L_F)$, it is clear that the effect of α on the shape of $\theta(L_F)$ becomes apparent only when $2\alpha > N/10$.

(b) *Specific binding in the presence of non-specific binding*

When there exists on a piece of DNA one specific site and N potential non-specific sites, Z truncated to degree 5 is given by:

$$Z = 1 + \sum_{i=1}^{i=5} \frac{\left(\prod_{j=1}^{j=5} [2(j-1)\alpha + N]\right)\left(\frac{K_S}{N} + K_N\right) K_N^{i-1} L_F^i}{i!}, \tag{11}$$

where K_S is the association constant for specific binding, K_N the association constant for non-specific binding, and α the co-operativity parameter that applies to interactions between all adjacent sites, including the specific site. If, however, the interactions between adjacent non-specific sites are described by a co-

operativity parameter ω_1, and that between the specific site and its adjacent non-specific sites by a co-operativity parameter, ω_2, Z truncated to degree 5, will be given by:

$$Z = 1 + K_S L_F + \sum_{i=2}^{i=5} \frac{\left(\prod_{j=2}^{j=5} |2\omega_2 + N + 2(j-2)\omega_1|\right) K_S K_N^{i-1} L_F^i}{i!} + \sum_{i=1}^{i=5} \frac{\left(\prod_{j=1}^{j=5} |2(j-1)\omega_1 + N|\right) K_N^i L_F^i}{i!}. \quad (12)$$

From equation (12) one can deduce that the effect of having $\omega_1 \neq \omega_2$ will become manifest only if $\omega_2 < \omega_1/2$ or $\omega_2 > 2\omega_1$.

For the special case where $\omega = 0$ and all the sites are non-overlapping, Z will be given exactly by the closed form expression:

$$Z = (1 + K_S L_F)(1 + K_N L_F)^N. \quad (13)$$

Equation (13) can also be used if the sites are overlapping and binding is non-co-operative, providing N is large ($N > 1000$) and $N \gg l$ (N is still taken as the total number of base-pairs per DNA molecule). From equations (1) and (13) a simple expression for K_S, which is useful in a preliminary assessment of equilibrium filter binding data, may be obtained in terms of K_N and L_{50} (the concentration of L_F at $\theta = 0{\cdot}5$):

$$K_S = \left[\frac{2}{(1 + K_N L_{50})^N} - 1\right] \frac{1}{L_{50}}. \quad (14)$$

For the special cases where $K_S \gg NK_N$ for equation (13), $K_S \gg NK_N$, ωK_N for equation (11) and $K_S \gg NK_N$, $\omega_1 K_N$, $\omega_2 K_N$ for equation (12), Z simplifies to $(1 + K_S L_F)$, and the corresponding equation for θ to a rectangular hyperbola; an example of this would be the binding of *lac* repressor to the *lac* operator as judged from the data of Riggs *et al.* (1970).

The effect of the introduction of a specific site in the presence of a large number of non-specific sites on equilibrium filter binding curves is shown in Figure 3 for the non-co-operative case ($\omega = 0$). For $K_S < NK_N/10$, the equilibrium filter binding curves are unaffected by the presence of the specific site and the curves will be identical to one with $K_S = 0$. For $K_S > NK_N/10$, the equilibrium filter binding curves become progressively less steep as K_S increases. When $K_S > 100\ NK_N$, the equilibrium filter binding curves are described by a rectangular hyperbola.

The extension of the theory to encompass more than a single specific site is easily achieved. For example, if we consider the non-co-operative case with $\omega = 0$ in a system with two specific sites with association constants K_{S1} and K_{S2}, and N non-specific sites with an association constant K_N, Z will be given exactly in the case of non-overlapping sites by:

$$Z = (1 + K_{S1} L_F)(1 + K_{S2} L_F)(1 + K_N L_F)^N. \quad (15)$$

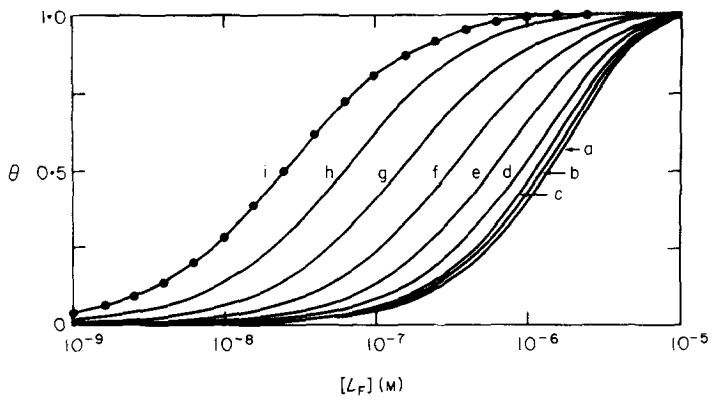

FIG. 3. The effect of a single specific site in the presence of N independent non-specific sites of equal intrinsic affinity on the shape of θ as a function of the free ligand concentration L_F for the case of non-co-operative binding.

The curves for $\theta(L_F)$ are calculated using eqns (1) and (11) for $K_N = 500\ M^{-1}$, $N = 1000$, $\alpha = 0$ and $K_S = 0\ M^{-1}$, $10^4\ M^{-1}$, $2{\cdot}51 \times 10^4\ M^{-1}$ (curve a), $6{\cdot}31 \times 10^4\ M^{-1}$ (curve b), $1{\cdot}58 \times 10^5\ M^{-1}$ (curve c), $3{\cdot}98 \times 10^5\ M^{-1}$ (curve d), $10^6\ M^{-1}$ (curve e), $2{\cdot}51 \times 10^6\ M^{-1}$ (curve f), $6{\cdot}31 \times 10^6\ M^{-1}$ (curve g), $1{\cdot}58 \times 10^7\ M^{-1}$ (curve h), and $3{\cdot}98 \times 10^7\ M^{-1}$ (curve i). A rectangular hyperbola calculated from $\theta = K_S L_F/(1 + K_S L_F)$ for $K_S = 3{\cdot}98 \times 10^7\ M^{-1}$ is also shown for comparison (●). Looking at the curves for $\theta(L_F)$, it is clear that the effect on the introduction of a specific site on the shape of $\theta(L_F)$ becomes apparent only when $K_S > NK_N/10$; when $K_S \gg NK_N$, $\theta(L_F)$ is described by a rectangular hyperbola (see curve i (●)).

3. Applications of the Theory of Equilibrium Filter Binding to Experimental Data

In this section we provide two examples of the analysis of experimental equilibrium filter binding data based upon the theoretical framework developed in the previous section, and discuss the implications of the additional insight thus obtained on these systems.

(a) *Specific and non-specific binding of the* Eco*RI restriction endonuclease to bacteriophage λ DNA*

Halford & Johnson (1980) examined the binding of the *Eco*RI restriction endonuclease to three derivatives of bacteriophage λ using the nitrocellulose filter binding assay. The three derivatives of bacteriophage λ were: λ395(0) (length 41 kb), from which all *Eco*RI recognition sites have been removed: λ401(2) (length 46 kb), which possesses only *sr*I 2 out of the five *Eco*RI recognition sites on wild-type λ DNA; and λ416(5) (length 41 kb), which carries only *sr*I 5 (Davidson & Szybalski, 1971; Murray & Murray, 1974). The equilibrium filter binding curves obtaining for the binding of *Eco*RI to λ401(2) and λ416(5) DNA were approximately hyperbolic. However, the curve for the binding of *Eco*RI to λ395(0) DNA exhibited a sigmoidal character and could be fitted to a second-order Adair equation, from which Halford & Johnson (1980) deduced that, whereas only one molecule of *Eco*RI binds to each specific site, two molecules of *Eco*RI bind to each non-specific site. This postulated mechanism of non-specific binding of *Eco*RI to DNA is not required to account for the data, however, and the marked deviation from

hyperbolic behaviour for the binding of *Eco*RI to λ395(0) DNA as monitored by the nitrocellulose filter binding assay is simply due to the fact that the form of the equilibrium filter binding curve for non-specific binding is given by equations (1) and (10). Moreover, a second-order Adair equation is an equation for $\theta_C(L_F)$ (defined by eqn (2)) in a two-site system, and not for $\theta(L_F)$ (defined by eqn (1)) in an N overlapping site system. Therefore, the use of a second-order Adair equation in fitting equilibrium filter binding data where the measured quantity is $\theta(L_F)$ is totally inappropriate, and the parameters obtained have no physical meaning.

In Figure 4, the data of Halford & Johnson (1980) are fitted simultaneously by non-linear optimization using equations (1) and (10) for the binding of *Eco*RI to λ395(0) DNA, and equations (1) and (11) for the binding of *Eco*RI to λ401(2) and λ416(5) DNA. Two calculations were performed: in both cases, the association constant K_N for non-specific binding of *Eco*RI to λ395(0), λ401(2) and λ416(5) DNA, the association K_{S1} for specific binding of *Eco*RI to λ401(2) DNA, and the association constant K_{S2} for specific binding of *Eco*RI to λ416(5) DNA were optimized; in one case, the co-operativity parameter ω was optimized, and in the other case, ω was set to zero and binding assumed to be non-co-operative. In both cases, the overall standard deviation of the fit is 3·5% and the distribution of residuals is random. The values and standard deviations of the optimized parameters for both cases are given in Table 1. With the exception of the co-operativity parameter ω, the parameters are well-determined with relative errors of less than 15%; ω, however, is determined only to within a factor of about 2. Moreover, the values obtained for K_N, K_{S1} and K_{S2} when ω is optimized are identical within the errors specified to those obtained when $\omega = 0$. We therefore

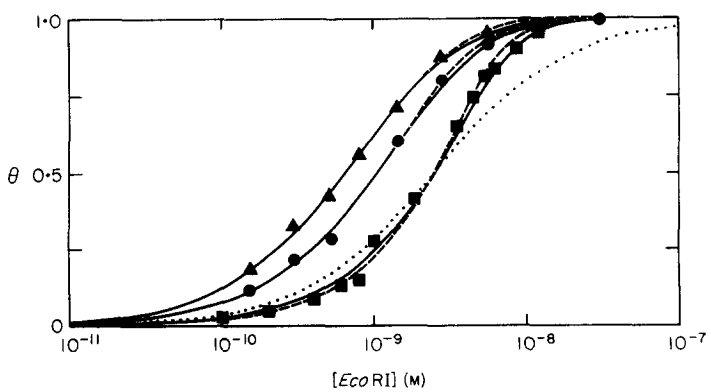

FIG. 4. Comparison of the experimental equilibrium nitrocellulose filter binding curves of Halford & Johnson (1980) for the binding of the *Eco*RI restriction endonuclease to λ395(0) (■), λ401(2) (▲) and λ416(5) (●) DNA with the theoretical equilibrium filter binding curves calculated using eqns (1), (10) and (11) with the best fit parameter values given in Table 1.

The theoretical curves for non-co-operative binding (i.e. $\omega = 0$) are shown as continuous lines and for co-operative binding as broken lines. Eqns (1) and (10) were used for the binding of *Eco*RI to λ395(0) DNA, which contains no specific recognition site, and eqns (1) and (11) for the binding of *Eco*RI to λ401(2) and λ416(5) DNA, which contain a single specific site, each in the presence of a large number of non-specific sites. All the experimental data were fitted simultaneously by non-linear optimization. A rectangular hyperbola given by $\theta_C = KL_F/(1+KL_F)$ for $K = 3{\cdot}98 \times 10^8\ M^{-1}$ is shown for comparison (dotted line).

TABLE 1

Equilibrium building parameters for the specific and non-specific interactions of the Eco*RI restriction endonuclease with DNA*

Parameter	Optimized value (±S.D.) Co-operative binding	Non-co-operative binding
$K_N(M^{-1})$	$6{\cdot}29 \times 10^3$ ($\pm 0{\cdot}5 \times 10^3$)	$6{\cdot}94 \times 10^3$ ($\pm 0{\cdot}3 \times 10^2$)
$K_{S1}(M^{-1})$	$1{\cdot}07 \times 10^9$ ($\pm 0{\cdot}1 \times 10^9$)	$1{\cdot}11 \times 10^9$ ($\pm 0{\cdot}1 \times 10^9$)
$K_{S2}(M^{-1})$	$5{\cdot}61 \times 10^8$ ($\pm 0{\cdot}7 \times 10^8$)	$5{\cdot}53 \times 10^8$ ($\pm 0{\cdot}7 \times 10^8$)
ω	$6{\cdot}69 \times 10^3$ ($\pm 5 \times 10^3$)	0
Overall S.D. fit (%)	3·6	3·5
$\bar{C}$†	0·74	0·48

Optimized values and standard deviations (S.D.) of the association constant K_N for non-specific binding of *Eco*RI to λ395(0) (41 kb), λ401(2) (46 kb) and λ416(5) (41 kb) DNA, the association constant K_{S1} for specific binding of *Eco*RI to λ401(2) DNA, the association constant K_{S2} for specific binding of *Eco*RI to λ416(5) DNA, and the co-operativity parameter ω, obtained by fitting all the experimental equilibrium filter binding data in Fig. 4 simultaneously using equations (1), (10) and (11). The overall S.D. values of the fits and the mean absolute correlation indices $\bar{C}$ are also given.

† The mean absolute correlation index $\bar{C}$ is given by:

$$\sum_{i=1}^{n} \left| \sum_{j=1}^{m} R_{ij} \Big/ \left(\sum_{j=1}^{m} R_{ij} \right)^{\frac{1}{2}} \right| \Big/ n,$$

where j identifies the independent variable (i.e. L_F) point and i the data curve, R_{ij} are the residuals, and n the number of curves. For $\bar{C} < 1{\cdot}0$, the distribution of residuals is random; for $\bar{C} \gg 1{\cdot}0$, there are systematic errors between observed and calculated curves (Clore & Chance, 1978*a*,*b*).

conclude that the data are inadequate to determine whether binding of *Eco*RI to DNA is co-operative or not. This will require obtaining data for the binding of *Eco*RI to DNA of lengths at least five to ten times shorter than that of λ395(0) DNA. It should also be noted that the equilibrium filter binding curves for the binding of *Eco*RI to λ401(2) and λ416(5) DNA deviate slightly from rectangular hyperbolas for $\theta > 0{\cdot}5$. Consequently, an estimate of the association constant K_{Si} for specific binding obtained by taking $K_{Si} \sim 1/L_{50}$ will be slightly larger than the true value of K_{Si}. The values of K_{S1} and K_{S2} obtained in this manner by Halford & Johnson (1980) are $1{\cdot}6 \times 10^9$ M^{-1} and $9{\cdot}1 \times 10^8$ M^{-1}, respectively, compared to the correct values of $1{\cdot}1 \times 10^9 \pm 0{\cdot}1 \times 10^9$ M^{-1} and $5{\cdot}6 \times 10^8 \pm 0{\cdot}7 \times 10^8$ M^{-1} given in Table 1. Nevertheless, these small differences do not affect the general conclusions of Halford & Johnson (1980) regarding specific binding of *Eco*RI to DNA.

(b) *Specific and non-specific binding of* L. casei *dihydrofolate reductase to pWDLcB1 and pBR322 DNA*

It has been demonstrated recently that *L. casei* dihydrofolate reductase has affinity for double-stranded DNA (Gronenborn & Davies, 1981). Gronenborn *et al.* (1981) examined the binding of DHFRase alone and in its binary and ternary complexes with folinic acid and NADPH to pBR322 DNA (length 4·36 kb: Sutcliffe, 1978) and pWDLcB1 DNA (length 7·26 kb), the latter differing from the former only in a 2·9 kb insert containing the DHFRase structural gene from a

methotrexate-resistant *L. casei* strain (Davies & Gronenborn, 1982), and found that the equilibrium binding curves obtained could be grouped into three sets (see Fig. 5).

(1) Those curves obtained for the binding of DHFRase alone or in the presence of folinic acid and/or NADPH (either as binary or ternary complexes) to pBR322 DNA that have a value of $L_{50} = 1{\cdot}8 \times 10^{-7}$ M.

(2) Those curves obtained for the binding of DHFRase alone and the DHFRase–NADPH binary complex to pWDLcB1 DNA that have a value of $L_{50} = 8 \times 10^{-8}$ M.

(3) Those curves obtained for the binding of DHFRase–folinic acid and the DHFRase–folinic acid–NADPH complexes to pWDLcB1 DNA that have a value of $L_{50} = 1{\cdot}2 \times 10^{-7}$ M.

Gronenborn *et al.* (1981) interpreted their data in a qualitative manner as follows: on the assumption that the binding of DHFRase alone and in its complexes with folinic acid and NADPH to pBR322 DNA only reflect non-specific binding, the observation that the corresponding filter binding curves are identical within experimental error, indicates that the affinity of the non-specific sites for DHFRase is unaffected by either folinic acid or NADPH; thus, the observations that the affinity of pWDLcB1 DNA for DHFRase alone and the DHFRase–NADPH complex is significantly greater than that for the complexes with folinic acid suggests that (1) DHFRase in the presence or absence of NADPH binds to a specific site on pWDLcB1 DNA at or near the DHFRase structural gene, and that (2) the affinity of the specific site for DHFRase is significantly reduced in the presence of folinic acid. Gronenborn *et al.* (1981) also suggested that the difference between the filter binding curves for the binding of the complexes of DHFRase with folinic acid to pWDLcB1 DNA and those for the binding of DHFRase alone and in its complexes with folinic acid and/or NADPH to pBR322 DNA could probably be accounted for by the slightly greater length of pWDLcB1 DNA relative to that of pBR322 DNA.

A preliminary assessment of the above hypothesis may be made by calculating the value of the apparent association constant K_{app} for the binding of DHFRase to DNA from the values of L_{50}, assuming non-co-operative binding using equation (6). The first set of curves for the binding of DHFRase and the DHFRase–NADPH, DHFRase–folinic acid, and DHFRase–folinic acid–NADPH complexes to pBR322 DNA yields a value of K_{app} of 880 M^{-1}: the second set of curves for the binding of DHFRase and the DHFRase–NADPH complex to pWDLcB1 DNA yields a value of K_{app} of 1200 M^{-1}: and the third set of curves for the binding of the DHFRase–folinic acid and DHFRase–folinic acid–NADPH complexes to pWDLcB1 DNA yields a value of K_{app} of 800 M^{-1}. Clearly, the value of K_{app} for the first and third sets of equilibrium filter binding curves are identical within experimental error. Thus, the difference between the first and third set of equilibrium filter binding curves can be accounted for purely on the basis of the different lengths of pBR322 and pWDLcB1 DNA. The value of K_{app}, however, for the second set of equilibrium filter binding curves is significantly different from that for the other two sets. This difference can be accounted for only by the presence of a site(s) of significantly higher affinity than that of the other non-specific sites. If we assume that only a single specific site of higher affinity is involved as proposed above, an estimate of

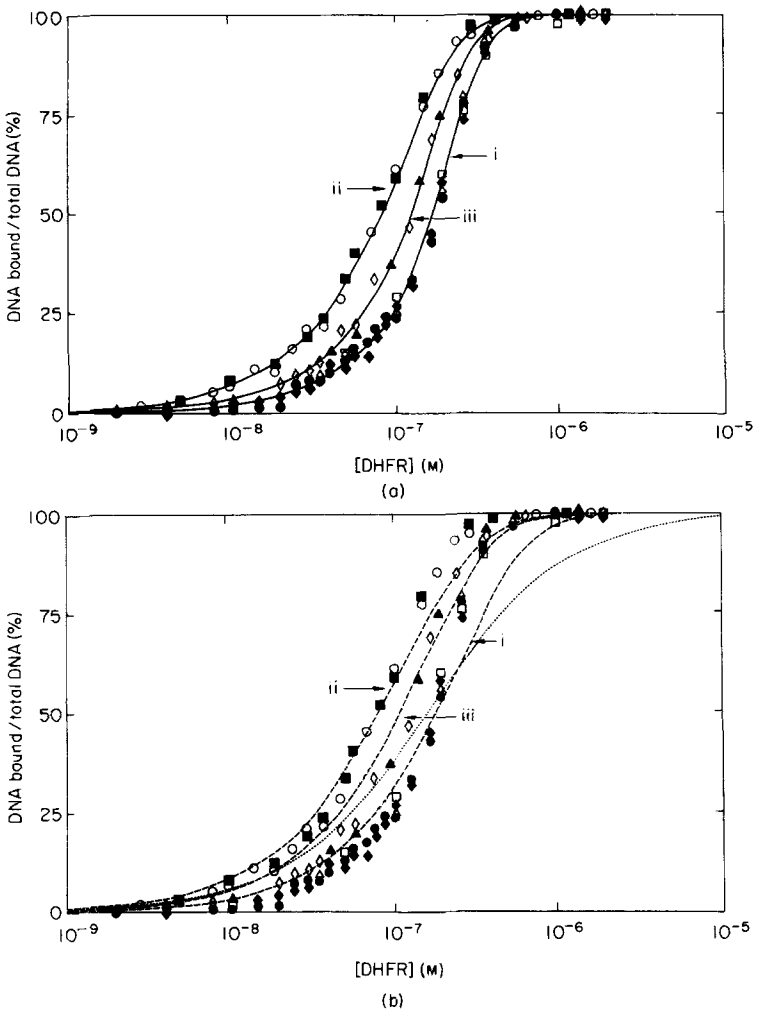

FIG. 5. Comparison of the experimental equilibrium filter binding curves determined by Gronenborn *et al.* (1981) for the binding of *L. casei* DHFRase to DNA from plasmids pBR322 and pWDLcB1, the latter differing from the former in a 2·9 kb insert containing the DHFRase structural gene from a methotrexate-resistant *L. casei* strain, with the theoretical equilibrium filter binding curves calculated using eqns (1), (10) and (11) with the best fit parameter values given in Table 1.

The experimental curves are shown as: curve i, binding of DHFRase alone (●) and the DHFRase–NADPH (□), DHFRase–folinic acid (△) and DHFRase–folinic acid–NADPH (◆) complexes to pBR322 DNA; curve ii, binding of DHFRase alone (○) and the DHFRase–NADPH complex (■) to pWDLcB1 DNA; curve iii, binding of the DHFRase–folinic acid (▲) and the DHFRase–folinic acid–NADPH (◇) complexes to pWDLcB1 DNA. The symbols are the means of the actual data points obtained from 3 separate experiments. The best fit theoretical curves obtained by fitting all the experimental data simultaneously using eqns (1), (10) and (11) with co-operative binding are shown in (a) as continuous lines and with non-co-operative binding (i.e. $\alpha = 0$) in (b) as broken lines. The values of the parameters used to calculate the best fit theoretical curves are given in Table 2. The best fit theoretical curves (curves i and iii) are calculated for non-specific binding of DHFRase to 4·36 and 7·26 kb lengths of DNA, respectively; curve ii is calculated for specific binding to a single site and non-specific binding of DHFRase to a 7·26 kb length of DNA. (b) A rectangular hyperbola given by $\theta_C = KL_F/(1 + KL_F)$ is shown for comparison with a value of $K = 5{\cdot}88 \times 10^6$ M^{-1} (dotted line).

the association constant K_{S1} for the binding of DHFRase alone and the DHFRase–NADPH complex to this site may be obtained using equation (14). Taking a value of 840 M^{-1} for K_N (the mean of the values of K_{app} obtained for the first and third sets of equilibrium filter binding curves), we obtain a value of 2.9×10^6 M^{-1} for K_{S1}.

A more rigorous assessment of the equilibrium filter binding data presented in Figure 5 requires a least-squares analysis in which all the experimental data are fitted simultaneously using equations (1), (10) and (11). Initially, we optimized four parameters: the association constant K_N for non-specific binding of DHFRase alone and in its ternary complexes with folinic acid and NADPH to pBR322 and pWDLcB1 DNA; the association constant K_{S1} for the binding of DHFRase alone and the DHFRase–NADPH complex to the specific site on pWDLcB1 DNA; the association constant K_{S2} for the binding of the DHFRase–folinic acid and the DHFRase–folinic acid–NADPH complexes to the specific site on pWDLcB1 DNA; and the co-operativity parameter, ω, for the interactions between adjacent sites. All the parameters were well-determined, with relative errors of less than 20%, except for the association constant K_{S2}, which was ill-determined. Thus, as expected from our preliminary assessment of the data, the equilibrium filter binding curves for the binding of DHFRase–folinic acid and DHFRase–folinic acid–NADPH complexes to pWDLcB1 DNA can be accounted for solely by the increased length of pWDLcB1 DNA relative to pBR322 DNA, indicating that $K_{S2} < N_{pWDLcB1} K_N/10$. In the subsequent optimization, all terms containing K_{S2} were neglected, and the parameters K_N, K_{S1} and ω were optimized. We also carried out one further calculation, in which ω was set equal to zero (i.e. no co-operativity) and K_N and K_{S1} were optimized. The overall standard deviation of the fits, the mean absolute correlation indices $\bar{C}$ (a measure of the distribution of residuals), and the values and standard deviations of the optimized parameters are given in Table 2. The best fit theoretical curves for co-operative and non-co-operative binding are shown in Figure 5(a) and (b), respectively. From both Figure 5 and Table 2, it is clear that the binding of DHFRase to DNA is a positively co-operative process, as there are clear systematic errors for the theoretical curves with $\omega = 0$. It is also interesting that the optimized values of K_{S1} for the co-operative and non-co-operative cases and the value of K_{S1} obtained in the preliminary calculation are all identical within a relative error of 10%.

The above data and analysis demonstrate the following four points.

(1) DHFRase alone and its binary and ternary complexes with folinic acid and NADPH bind non-specifically to DNA with an association constant K_N of 509 M^{-1}.

(2) DHFRase alone and the DHFRase–NADPH complex bind to a specific site at or near the DHFRase structural gene on pWDLcB1 DNA with an association constant K_{S1} of 3.66×10^6 M^{-1}. The ratio of the association constant for specific (K_{S1}) to non-specific (K_N) binding is of the order of 7×10^3, corresponding to a difference in the free energy of binding of about 22 kJ mol^{-1}.

(3) In the presence of folinic acid, the association constant for specific binding is reduced to a value below $N_{pWDLcB1} \cdot K_N/10 \sim 3.7 \times 10^5$ M^{-1}, such that the equilibrium filter binding for the binding of DHFRase–folinic acid and DHFRase–folinic acid–NADPH complexes to pWDLcB1 DNA can be accounted for simply by the increased length of pWDLcB1 DNA relative to pBR322 DNA.

TABLE 2

Equilibrium binding parameters for the specific and non-specific interactions of L. casei *dihydrofolate reductase with DNA*

	Co-operative binding	Non-co-operative binding ($\alpha = 0$)
$K_{S1}(M^{-1})$†	$3{\cdot}66 \times 10^6$ ($\pm 0{\cdot}4 \times 10^6$)	$3{\cdot}30 \times 10^6$ ($\pm 0{\cdot}5 \times 10^6$)
$K_N(M^{-1})$	$5{\cdot}09 \times 10^2$ ($\pm 0{\cdot}3 \times 10^2$)	$8{\cdot}47 \times 10^2$ ($\pm 0{\cdot}3 \times 10^2$)
α‡	$5{\cdot}00 \times 10^3$ ($\pm 0{\cdot}5 \times 10^3$)	0
Overall S.D. of fit (%)§	5·8	6·9
$\bar{C}$‖	0·46	1·8

Values of the optimized parameters, overall standard deviation (S.D.) of the fits and mean absolute correlation indices ($\bar{C}$), obtained by fitting all the experimental equilibrium filter binding data for the binding of *L. casei* DHFRase to pBR322 (4·36 kb) and pWDLcB1 (7·26 kb) DNA in Fig. 5 simultaneously using equations (1), (10) and (11). (The S.D. values of the optimized parameters are shown in parentheses.)

† K_{S1} is the association constant for specific binding of DHFRase alone and the DHFRase–NADPH complex to a site at or near the DHFRase structural gene (from a methotrexate-resistant strain of *L. casei*) in the plasmid pWDLcB1. K_N is the association constant for non-specific binding of DHFRase alone and the DHFRase–NADPH, DHFRase–folinic acid and DHFRase–folinic acid–NADPH complexes to DNA.

‡ It will be noted that the co-operativity parameter α applies to all interactions involving adjacent sites, including the specific site. In our preliminary optimizations, we used 2 separate co-operativity parameters, α_1 to describe interactions between adjacent and non-specific sites, and α_2 for interactions between the specific site and its adjacent non-specific sites (see eqn (12)). However, although the optimized value of α_2 was approximately 50% greater than that of α_1, α_2 was determined only to within a factor of 10. We therefore concluded that the experimental data could not discriminate between these interactions. Thus, in all subsequent calculations, we made no distinction between these interactions and used only a single co-operativity parameter, α, to describe them.

§ The overall standard error of the data is $6{\cdot}0 \pm 0{\cdot}5\%$.

‖ The mean absolute correlation index $\bar{C}$ is defined in the footnote to Table 1.

(4) Binding of DHFRase alone and its binary and ternary complexes with folinic acid and NADPH to DNA is positively co-operative, involving interactions between adjacent sites (including the specific site). The optimized value of the co-operativity parameter α is $5{\cdot}0 \times 10^3$. Using equation (9), one can estimate that the value for the local co-operativity parameter β lies between 330 and 500 on the assumption that co-operativity extends only to the two contiguous sites on either side of the specific site, and that the length of each binding site for DHFRase on DNA lies in the range of 10 to 15 base-pairs (which would be expected on the basis of the DHFRase molecular dimensions of 37 Å × 30 Å × 45 Å; D. A. Matthews, personal communication). This range of values for β lies within the range of values for the co-operativity parameter ω for contiguous binding, as defined by McGhee & von Hippel (1974), reported for other DNA binding proteins (Saxe & Revzin, 1979; Takahashi *et al.*, 1979; Newport *et al.*, 1981).

(c) *General comments*

A particularly important point in the analysis of equilibrium binding data is to fit the data to a physically meaningful model. For example, if one were to measure the number of moles of protein bound per mole of DNA, $\theta_C(L_F)$, as a function of the

free protein concentration L_F, in a system with only two non-overlapping binding sites on the DNA, $\theta_C(L_F)$ would be given appropriately by a second-order Adair equation. If, on the other hand, one were to measure the fraction, θ, of DNA to which at least one protein molecule is bound as a function of L_F, in a system with N overlapping non-specific sites on the DNA, it would be totally inappropriate to use a second or higher-order Adair equation to fit the data, even though they may be able to fit the data perfectly satisfactorily, as the parameters obtained would be physically meaningless. Moreover, if one considered not a single equilibrium binding curve of θ *versus* L_F for this system, but a set of such equilibrium binding curves obtained using the same protein but different lengths of DNA, all the equilibrium binding curves could be fitted simultaneously using the appropriate equation for $\theta(L_F)$ (eqns (1) and (10)) using a single parameter if non-specific binding is non-co-operative and two parameters if non-specific binding is co-operative. If, however, a second or higher-order Adair equation was used to fit the same data set, each equilibrium binding curve could be fitted only individually, and two or more parameters would be required per curve. This is illustrated by the two examples in Figures 4 and 5. In both cases, each equilibrium binding curve can indeed be fitted individually by a second-order Adair equation yielding two physically meaningless parameters per curve. However, using the appropriate equations for $\theta(L_F)$ (eqns (1) and (10) for non-specific binding and eqns (1) and (11) for specific binding), only three parameters are required to fit the three equilibrium binding curves in Figure 4, namely: the association constant K_N for non-specific binding of *Eco*RI to λ395(0), λ401(2) and λ416(5) DNA, the association constant K_{S1} for specific binding of *Eco*RI to λ401(2) DNA and the association constant K_{S2} for specific binding of *Eco*RI to λ416(5) DNA. Similarly, in the case of the three sets of equilibrium binding curves in Figure 5, only three parameters are required to fit all the data simultaneously: namely, the association constant K_N for non-specific binding of DHFRase alone and in its complexes with folinic acid and/or NADPH to DNA from the two plasmids pBR322 and pWDLcB1, the association constant K_{S1} for specific binding of DHFRase alone and in its complex with NADPH to a specific site on pWDLcB1 DNA, and the co-operativity parameter α.

A second important point in the analysis of equilibrium binding curves for θ *versus* L_F concerns the approach required to detect specific binding. It will be noted, looking at Figure 3, that the effect of introducing a specific site on the shape of $\theta(L_F)$ is simply to reduce its steepness. As a result, it will usually be possible to fit such a curve using equations (1) and (10) for non-specific binding by optimizing both the association constant K_N for non-specific binding and the co-operativity parameter α (which may be negative). It is therefore absolutely essential to obtain data for θ *versus* L_F using DNA where there is no specific binding site and only non-specific binding occurs, and comparing this equilibrium binding curve with that obtained using DNA with the putative specific site. If the two equilibrium binding curves cannot be fitted using equations (1) and (10) for non-specific binding with a single set of two parameters, K_N and α, then the presence of a specific site can be inferred. The two equilibrium binding curves should then be fitted simultaneously using equations (1) and (10) for the curve where only non-specific binding occurs, and equations (1) and (11) for the curve where both specific and non-specific

binding occur, by optimizing three parameters, K_N, λ and the association constant K_S for specific binding. The advantage of fitting both equilibrium binding curves simultaneously is that they both contain information on non-specific binding, so that the parameters describing non-specific binding, K_N and λ, can be better determined. If the equilibrium binding curves were fitted individually, it would be necessary first to obtain values of K_N and λ by fitting the equilibrium binding curve where only non-specific binding is involved using equations (1) and (10). A value for K_S would then be obtained by fitting the equilibrium binding curve involving specific and non-specific binding using equations (1) and (11), and optimizing K_S while holding K_N and λ constant at their values obtained previously. This is because the equilibrium binding curve involving both specific and non-specific binding cannot by itself determine all three parameters K_S, K_N and λ, whereas the two equilibrium binding curves taken together can.

4. Discussion

In this paper we have developed the theoretical framework required for the analysis of equilibrium binding data obtained by any method that monitors the fraction, θ, of DNA to which at least one protein molecule is bound, as a function of the free protein concentration L_F. The two methods that measure $\theta(L_F)$ are the well-known nitrocellulose filter binding assay (Riggs *et al.*, 1970; Hinkle & Chamberlin, 1972) and the more recent immunoprecipitation assay (McKay, 1981). The theoretical framework is simple, and expressions for $\theta(L_F)$ are very easy to derive and evaluate. The cases considered explicitly in this paper for which expressions for $\theta(L_F)$ are given, are the co-operative and non-co-operative equilibrium binding of a protein to a large number of non-specific sites and to a specific site(s) in the presence of a large number of non-specific sites. The ease with which the theory is applied to the analysis of experimental nitrocellulose filter binding data is illustrated in section 3 of this paper for the binding of the *Eco*RI restriction endonuclease both non-specifically and to specific recognition sites on bacteriophage λ DNA, and for the co-operative binding of the enzyme dihydrofolate reductase both non-specifically to DNA and to a specific site at or near its own structural gene.

A number of questions, both of a theoretical and a practical nature, concerning the use of methods that measure the quantity θ as a function of the free protein concentration L_F, compared to methods that measure the number of moles of protein bound per mole of DNA, $\theta_C(L_F)$ as a function of L_F, immediately come to mind, and are considered below.

(a) *Non-specific binding*

Most studies in the literature concerning non-specific binding of proteins to DNA have used methods that measure the quantity $\theta_C(L_F)$ (see e.g. Revzin & von Hippel, 1977; Butler *et al.*, 1977; de Haseth *et al.*, 1977; Saxe & Revzin, 1979; Takahashi *et al.*, 1979; Newport *et al.*, 1981). Analysis of such data requires the use of one of the

formalisms to deal with the problem of overlapping sites (McGhee & von Hippel, 1974; Schellman, 1974; Schwarz, 1977; Epstein, 1978). As a consequence, three major theoretical disadvantages of measuring $\theta_C(L_F)$ emerge.

(1) In all the formalisms developed, θ_C is an implicit variable appearing on both the right and left-hand side of the equations (e.g. see eqn (15) of McGhee & von Hippel, 1974). As a result, a direct solution of $\theta_C(L_F)$ as a function of L_F requires an iterative method, which is necessarily complex. The procedure that has been used in the literature to circumvent this problem is one in which θ_C is treated as an independent variable and the quantity θ_C/L_F calculated. That is to say, the data are manipulated into the Scatchard plot formalism. The disadvantages of such an approach are obvious; namely, the introduction of large errors on both the abscissa and ordinate. Moreover, as regards least-squares fitting procedures, the Scatchard treatment is strictly speaking illegitimate, because θ_C is not an independent variable but an experimental dependent variable, which varies as a function of the independent variable L_F.

(2) The equations for θ_C/L_F contain three unknowns: the association constant K, the site size l, and the co-operativity parameter ω. However, because of the overlapping site problem, l and ω are necessarily correlated, and consequently it is difficult to determine all three parameters accurately.

(3) The treatment of the co-operativity parameter ω is model dependent. Thus, in the formalism of McGhee & von Hippel (1974), ω applies only to interactions between contiguous sites.

The theoretical advantages of measuring the quantity $\theta(L_F)$ over measuring the quantity $\theta_C(L_F)$ now become obvious.

(1) For large values of $N(> 1000)$ and $N \gg l$, the overlap problem is completely avoided because θ will have a value of $> 0{\cdot}99$ for very small values of θ_C (≥ 5). Consequently, θ may be calculated explicitly in terms of L_F and no data manipulation of the Scatchard type is required.

(2) The equations for θ (see eqns (1) and (10)) contain only two unknowns: the association constant K and the co-operativity parameter α.

(3) The co-operativity parameter α is essentially model independent and is also independent of the site size l. From α, a local co-operativity parameter β may be evaluated (see eqn (9)) if the distance m in base-pairs over which the co-operative process operates is known by independent means. m may not necessarily coincide with the site size l. If, however, one assumes that $m = l$, the co-operativity parameter β is equivalent to the co-operativity parameter ω of McGhee & von Hippel (1974).

The disadvantage in measuring $\theta(L_F)$ is that the site size l cannot be determined.

From a practical point of view, there are no particular advantages or disadvantages in measuring $\theta(L_F)$ rather than $\theta_C(L_F)$, as the methods for measuring both $\theta(L_F)$ and $\theta_C(L_F)$ are both relatively easy. Suffice it to say that if a distinct change in the optical absorption or fluorescence properties of the protein under consideration occur on binding to DNA, then the measurement of $\theta_C(L_F)$ is significantly faster than that of $\theta(L_F)$. However, if such changes do not occur, and one has to resort either to equilibrium sedimentation (Jensen & von Hippel, 1977; Revzin & von Hippel, 1977) or gel filtration (de Haseth *et al.*, 1977) methods to

measure $\theta_C(L_F)$, then the measurement of $\theta(L_F)$ is significantly faster than that of $\theta_C(L_F)$.

(b) *Specific binding in the presence of non-specific binding*

If binding is non-co-operative, θ_C for the binding of a protein to x independent specific sites with association constants K_{Si} in the presence of N independent non-specific sites of equal intrinsic affinity, each with an association constant K_N, will be given by:

$$\theta_C = \sum_{i=1}^{i=x} \left(\frac{K_{Si} L_F}{1 + K_{Si}} \right) + N K_N L_F \tag{16}$$

at low saturation levels under conditions where N is large ($N > 1000$) and $N \gg l$. From equation (16) it is clear that providing $K_{Si} > NK_N/10$, binding to the ith specific site can be detected, providing an experimental method of very high sensitivity is used. This invariably involves a method in which the protein is radioactively labelled. Analogous methods for looking at the binding of small radiolabelled agonists to receptors in subcellular fractions have been used with great success (Birdsall & Hulme, 1976; Hulme *et al.*, 1978; Birdsall *et al.*, 1980). However, in these cases, non-specific binding of the radiolabelled ligand is invariably non-co-operative. In the case of protein–DNA interactions, however, non-specific binding is frequently co-operative (Takahashi *et al.*, 1979; Saxe & Revzin, 1979; Newport *et al.*, 1981; and section 3 of this paper) and, in such cases, equation (16) will only hold if, in the formalism developed in this paper, $K_S \gg 2\omega K_N$. If this condition is not fulfilled, the equations for $\theta_C(L_F)$ become exceedingly complicated, even when appropriate approximations are made for conditions of low saturation ($\theta_C < 5$), making the analysis of experimental data extremely difficult. In contrast, the introduction of a specific site(s) in the equations for $\theta(L_F)$ is easy to derive and evaluate whether binding is co-operative or not, and does not in any way complicate the analysis of the data.

In addition to the above theoretical advantage, the measurement of $\theta(L_F)$ also carries an experimental advantage. The measurement of $\theta(L_F)$ only requires radioactive end-labelling of DNA pieces with ^{32}P; this can cause only a very minor perturbation in the structure and properties of the DNA. The measurement of $\theta_C(L_F)$, however, requires radioactive labelling of the protein, which could potentially cause a more extensive perturbation in the structure and properties of the protein and, in particular, may affect the DNA binding site on the protein.

We thank Sir Arnold Burgen and Dr G. C. K. Roberts for helpful discussions.

REFERENCES

Bailey, J. M. (1979). *Anal. Biochem.* **93**, 204–206.
Birdsall, N. J. M. & Hulme, E. C. (1976). *J. Neurochem.* **27**, 7–16.
Birdsall, N. J. M., Hulme, E. C. & Burgen, A. S. V. (1980). *Proc. Roy. Soc. Ser. B*, **207**, 1–12.
Butler, A. P., Revzin, A. & von Hippel, P. H. (1977). *Biochemistry*, **16**, 4757–4768.

Cantor, C. R. & Schimmel, P. R. (1980). *Biophysical Chemistry*, vol. 3, W. H. Freeman and Co., San Francisco.
Clore, G. M. & Chance, E. M. (1978*a*). *Biochem. J.* **173**, 799–810.
Clore, G. M. & Chance, E. M. (1978*b*). *Biochem. J.* **175**, 709–725.
Davidson, N. & Szybalski, W. (1971). In *The Bacteriophage λ* (Hershey, A. D., ed.), pp. 45–82, Cold Spring Harbor Laboratory Press, Cold Spring Harbor.
Davies, R. W. & Gronenborn, A. M. (1982). *Gene*. In the press.
de Haseth, P. L., Gross, C. A., Burgess, R. R. & Record, T. M. (1977). *Biochemistry*, **16**, 4777–4783.
Epstein, I. R. (1978). *Biophys. Chem.* **8**, 327–339.
Giacomini, P. U. (1976). *FEBS Letters*, **72**, 83–86.
Giacomini, P. U. (1979). *Eur. J. Biochem.* **98**, 557–566.
Gronenborn, A. M. & Davies, R. W. (1981). *J. Biol. Chem.* **256**, 12152–12157.
Gronenborn, A. M., Clore, G. M. & Davies, R. W. (1981). *FEBS Letters*, **133**, 92–94.
Halford, S. E. & Johnson, N. P. (1980). *Biochem. J.* **191**, 593–604.
Hinkle, D. C. & Chamberlin, M. J. (1972). *J. Mol. Biol.* **70**, 157–185.
Hulme, E. C., Birdsall, N. J. M., Burgen, A. S. V. & Mehta, P. (1978). *Mol. Pharmacol.* **14**, 737–750.
Jensen, D. E. & von Hippel, P. H. (1977). *Anal. Biochem.* **80**, 267–281.
McGhee, J. D. & von Hippel, P. H. (1974). *J. Mol. Biol.* **86**, 469–489.
McKay, R. D. G. (1981). *J. Mol. Biol.* **145**, 471–488.
Murray, K. & Murray, N. E. (1974). *Nature (London)*, **251**, 476–481.
Newport, J. W., Lonberry, N., Kowalczykowski, S. C. & von Hippel, P. H. (1981). *J. Mol. Biol.* **145**, 105–121.
Poland, D. (1978). *Co-operative Equilibria in Physical Biochemistry*, Clarendon Press, Oxford.
Revzin, A. & von Hippel, P. H. (1977). *Biochemistry*, **16**, 4769–4776.
Riggs, A. D., Suzuki, H. & Bourgeois, S. (1970). *J. Mol. Biol.* **48**, 67–83.
Saxe, S. A. & Revzin, A. (1979). *Biochemistry*, **18**, 255–263.
Schellman, J. A. (1974). *Israel J. Chem.* **12**, 219–238.
Schwarz, G. (1977). *Biophys. Chem.* **6**, 65–76.
Strauss, H. S., Burgess, R. R. & Record, M. T. (1980). *Biochemistry*, **19**, 3504–3515.
Sutcliffe, J. G. (1978). *Cold Spring Harbor Symp. Quant. Biol.* **43**, 77–90.
Takahashi, M., Blazy, B. & Baudras, A. (1979). *Nucl. Acids Res.* **7**, 1699–1712.
von Hippel, P. M. & McGhee, J. D. (1972). *Annu. Rev. Biochem.* **41**, 231–300.
Wyman, J. (1967). *J. Amer. Chem. Soc.* **89**, 2202–2218.

Edited by G. A. Gilbert